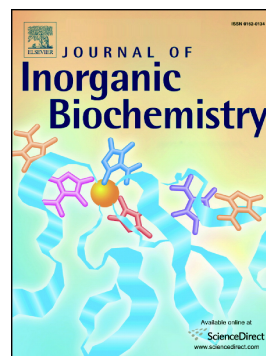


Journal Pre-proof

Unfolding cytochromes c-b562 and Rd apo b562

John J. Kozak, Harry B. Gray, Roberto A. Garza-López



PII: S0162-0134(20)30237-3

DOI: <https://doi.org/10.1016/j.jinorgbio.2020.111209>

Reference: JIB 111209

To appear in: *Journal of Inorganic Biochemistry*

Received date: 30 April 2020

Revised date: 26 July 2020

Accepted date: 28 July 2020

Please cite this article as: J.J. Kozak, H.B. Gray and R.A. Garza-López, Unfolding cytochromes c-b562 and Rd apo b562, *Journal of Inorganic Biochemistry* (2020), <https://doi.org/10.1016/j.jinorgbio.2020.111209>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

Unfolding Cytochromes c-b₅₆₂ and Rd apo b₅₆₂

John J. Kozak ^a, Harry B. Gray ^b and Roberto A. Garza-López ^c

a) Department of Chemistry, DePaul University, Chicago IL 60604-6116.

b) Beckman Institute, California Institute of Technology, Pasadena, CA 91125

c) Department of Chemistry and Seaver Chemistry Laboratory,
Pomona College, Claremont, CA 91711

ABSTRACT

We have analyzed the early stages of unfolding of cytochromes c-b₅₆₂ (PDB ID: 2BC5) and Rd apo b₅₆₂ (PDB ID: 1YYJ). Our geometric approach proceeds from an analysis of the crystal structure reported for each protein. We quantify, residue-by-residue and region-by-region, the spatial and angular changes in the structure as the protein denatures, and quantify differences that result from the seven residues that differ in the two proteins. Using two independent analyses, one based on spatial metrics and the second on angular metrics, we establish the order of unfolding of the five helices in cyt c-b₅₆₂ and the four helices in the apo protein. For the two helices nearest the N-terminal end of both proteins, the ones in the apo protein unfold first. For the two helices nearest the C-terminal end, the interior helix of the apo protein unfolds first, whereas the terminal helix of the holo protein unfolds first. Excluded-volume effects (repulsive interactions) are minimized in turning regions; the overall range in Δ values is $\Delta=36.3 \text{ \AA}^3$ for cyt c-b₅₆₂ and $\Delta=36.6 \text{ \AA}^3$ for the apo protein, whereas the span for all 20 amino acids is $\Delta=167.7 \text{ \AA}^3$. As our work indicates that the interior helix of cytochrome c-b₅₆₂ is the first to fold, we suggest that this helix protects the heme from misligation, consistent with ultrafast folding over a minimally frustrated funneled landscape.

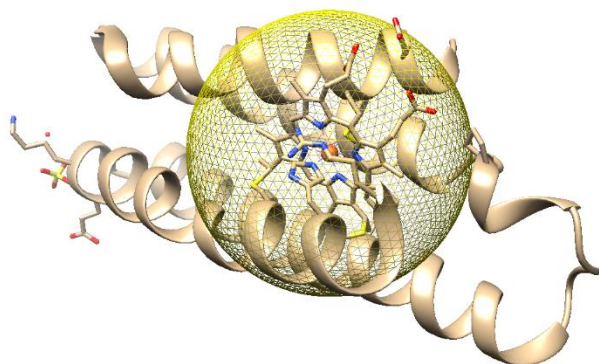
I. Introduction

The pioneering investigations of Wolynes, Saven, and Onuchic [1-5] on funneled energy landscapes as well as definitive experimental work by Dobson, Eaton, and others [6-9] have greatly advanced our understanding of protein folding. Additional insights have followed from research on helical proteins by Thirumalai [10-13] using force fields that depend on the radial and angular orientation of residues to recognize native conformations.

Like Thirumalai, we have been interested in the angle landscapes of helical proteins [14-16]. But our methodology is very different, as we have employed a geometrical approach [14] to analyze these landscapes. We have been particularly focused on heme protein landscapes [9,17-19]. Why are they so different? Why does cyt c-b₅₆₂ fold so rapidly; and more specifically, how is heme misligation avoided as the protein rocks down a minimally frustrated landscape [9]. We have looked more deeply into these landscapes, by characterizing and distinguishing, residue-by-residue and region-by-region, the response of each cytochrome as it unfolds from its native state.

Crystal structures have been reported for cyt c-b₅₆₂ [19] and Rd apo cyt b₅₆₂ [20], providing data for an in-depth study of the respective roles of the metal ion and the hydrophobic region in influencing the unfolding of both proteins. We use the spatial coordinates of cyt c-b₅₆₂ from *E. coli* [19] and those of a redesigned cytochrome, Rd apo b₅₆₂ [20]: cytochrome c-b₅₆₂ has 106 residues, with five helices and a heme; we picked Rd apo b₅₆₂ for comparison, because it also has 106 residues (even though it has an extended turning region and one fewer helix). A principal objective of our study is to characterize the spherical region in the cyt c-b₅₆₂ structure encompassed by residues within a 10 Å radius of the heme (Figure 1).

Figure 1. Structure of cyt c-b₅₆₂ highlighting residues within 10 Å of the heme.



II. Signatures of Unfolding

For each residue i , the coordinates and distance R_i from the crystallographic origin can be determined. Any point can be chosen as the center of a coordinate system, and here we choose the cyt c-b₅₆₂ Fe atom. Numerical values of metric, calculated assuming different origins will, of course, be different, but metrics that depend on the relative difference between two distances, *e.g.* $R_2 - R_1$, will be the same. For the apo protein, we choose the origin to be the Fe site in cyt c-b₅₆₂ (Figure 2).

With respect to this common origin, the distance of each residue i is given in Figure 3. The profile for cyt c-b₅₆₂ is displayed in the upper figure, and for the apo protein in the lower one. In both figures, the helix nearest the N-terminal end of each protein is in magenta, the neighboring helix in red. In both figures, the helix closest to the C-terminal end is color coded in cyan, and the neighboring helical region in blue. In cyt c-b₅₆₂, the (fifth) helix between the two N-terminal helices and the two C-terminal helices is in gold. All turning regions are in green. Noticeable is the extended turning (hydrophobic) region, residues 42 to 57, in the apo protein. The magnitude of R_i will depend on the choice of origin; as a result, the vertical scale for the two proteins is different.

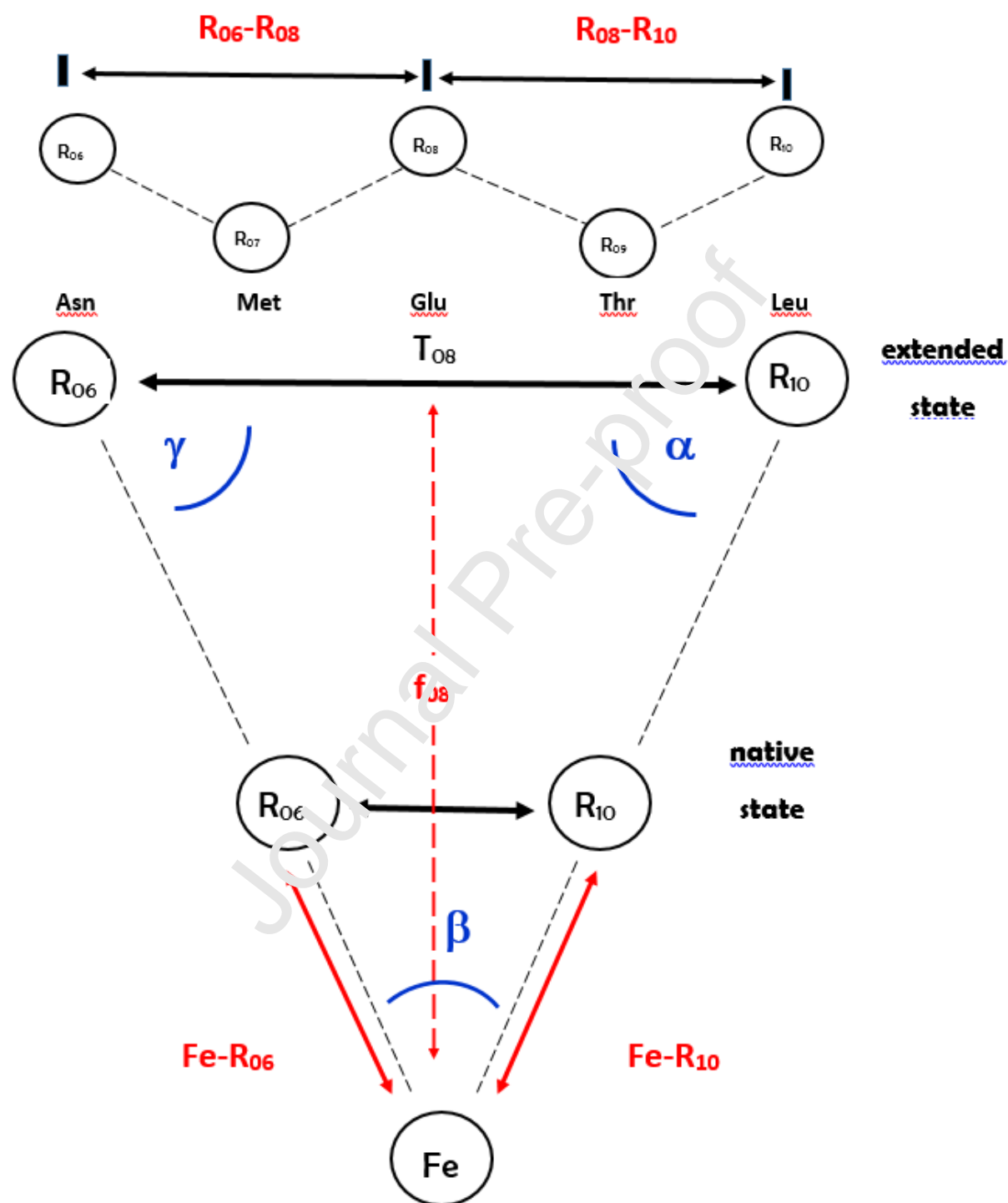
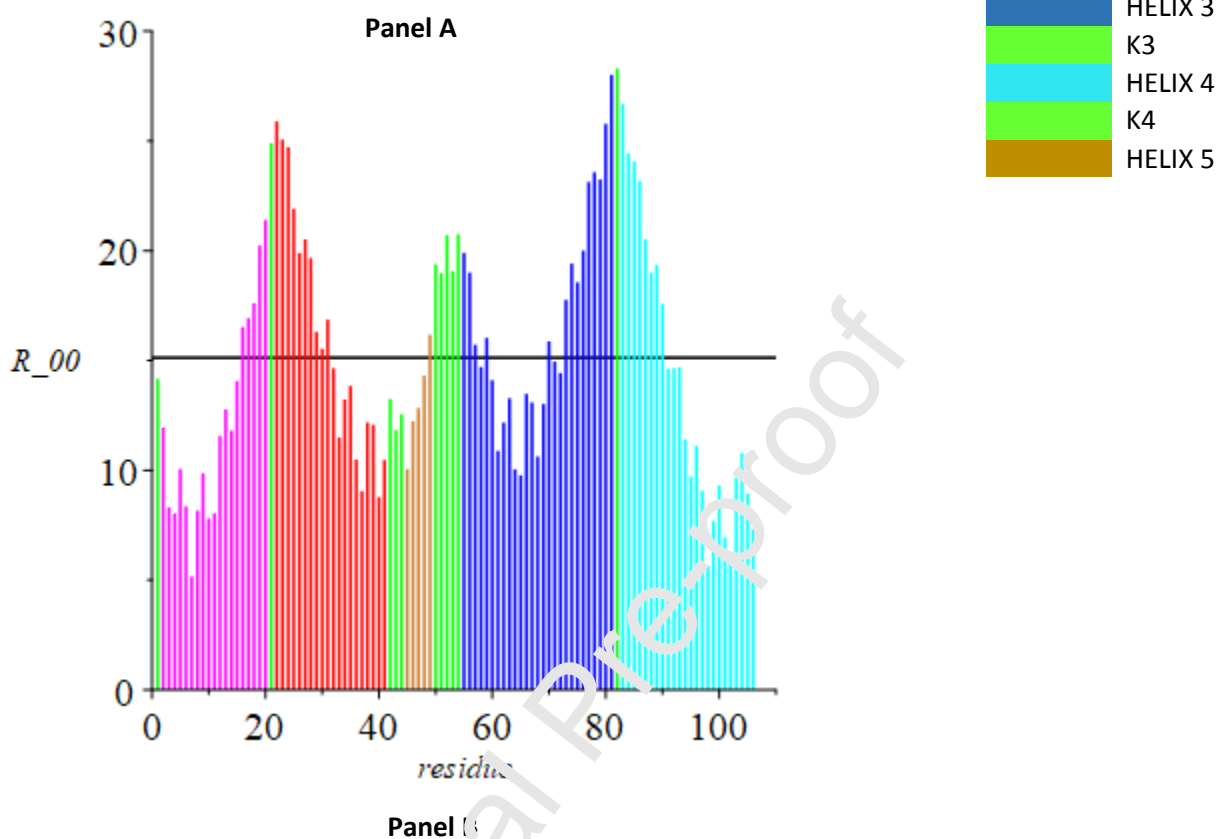
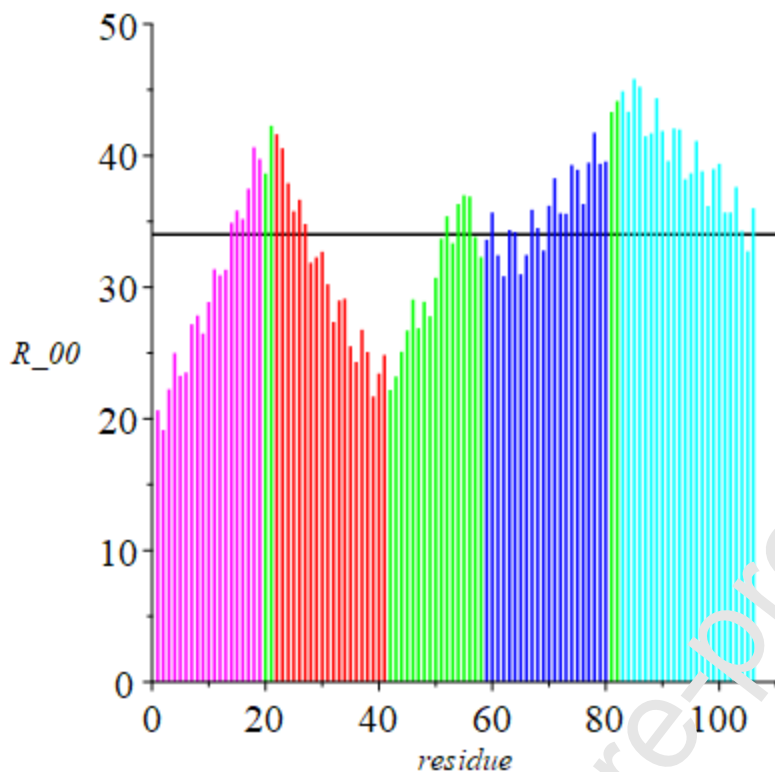
Figure 2: Geometrical model for a five-residue segment in cytochrome c-b₅₆₂.

Figure 3. Distance (\AA) of residue i from Fe. Upper /lower figure: cyt c-
b₅₆₂/ Rd apo cyt b₅₆₂.





We now introduce two metrics that depend only on the relative distance between two values R_i ; these metrics will have the same calculated values irrespective of the choice of origin. We begin by choosing a triplet T_i centered on residue i , viz., residues $R(i-1)$, $R(i)$ and $R(i+1)$, as our modular unit. Since the crystal structure represents the globally optimized geometry of the native state, this triad captures (exactly) the attractive and repulsive (steric) interactions between and among the three residues in the native state and, once determined from crystal structure data, remains invariant in our approach.

The geometry of a segment of five residues in the first stage of unfolding can be represented by annexing two triads in a *planar* configuration. For example, for a five-residue segment centered on residue $i=8$, the linear distance T_{08} between the α -carbons in the terminal residues ($i=6$ and $i=10$) of this segment is: $T_{08} = R_{06to08} + R_{08to10}$. A segment of seven residues centered on

residue $i=8$ in the second stage of unfolding is similarly described as an annexation of three triplets; $T08 = R05\text{to}07 + R07\text{to}09 + R09\text{to}11$. The longest array of triplets considered in this study is a segment of fifteen residues; for $i=8$ the annexation of seven triplets in a planar configuration is given by

$$T08 = R01\text{to}03 + R03\text{to}05 + R05\text{to}07 + R07\text{to}09 + R09\text{to}11 + R11\text{to}13 + R13\text{to}15.$$

In the native state, distances between the α -carbons in the terminal residues of each of the above segments can be calculated from crystallographic data. Specifically, in the above example, we have for the five-residue segment centered on residue $i=8$,

$$R06\text{to}10 = \sqrt{(R10x - R6x)^2 + (R10y - R6y)^2 + (R10z - R6z)^2},$$

for the seven-residue segment,

$$R05\text{to}11 = \sqrt{(R11x - R5x)^2 + (R11y - R5y)^2 + (R11z - R5z)^2}, \text{ and for the fifteen-residue segment,}$$

$$R01\text{to}15 = \sqrt{(R15x - R1x)^2 + (R15y - R1y)^2 + (R15z - R1z)^2}.$$

We next introduce two metrics using the above defined distances. The first metric is the difference δ in distance (\AA) of (here) a five-residue *linear* extension of triplets *minus* the crystallographic distance between terminal α -carbons,

$$\delta_{08} = T08 - R06\text{to}10.$$

Values of $\delta \sim 0$ identify five-residue segments whose geometry is essentially that of the unfolded state, with values $\delta > 1$ reflecting the persistence of the native-state geometry. Thus, for a given helical region, the larger the value of δ , the more the native-state character is conserved.

Complementary to the metric δ is the *ratio* of a n -residue *linear* extension of triplets to the

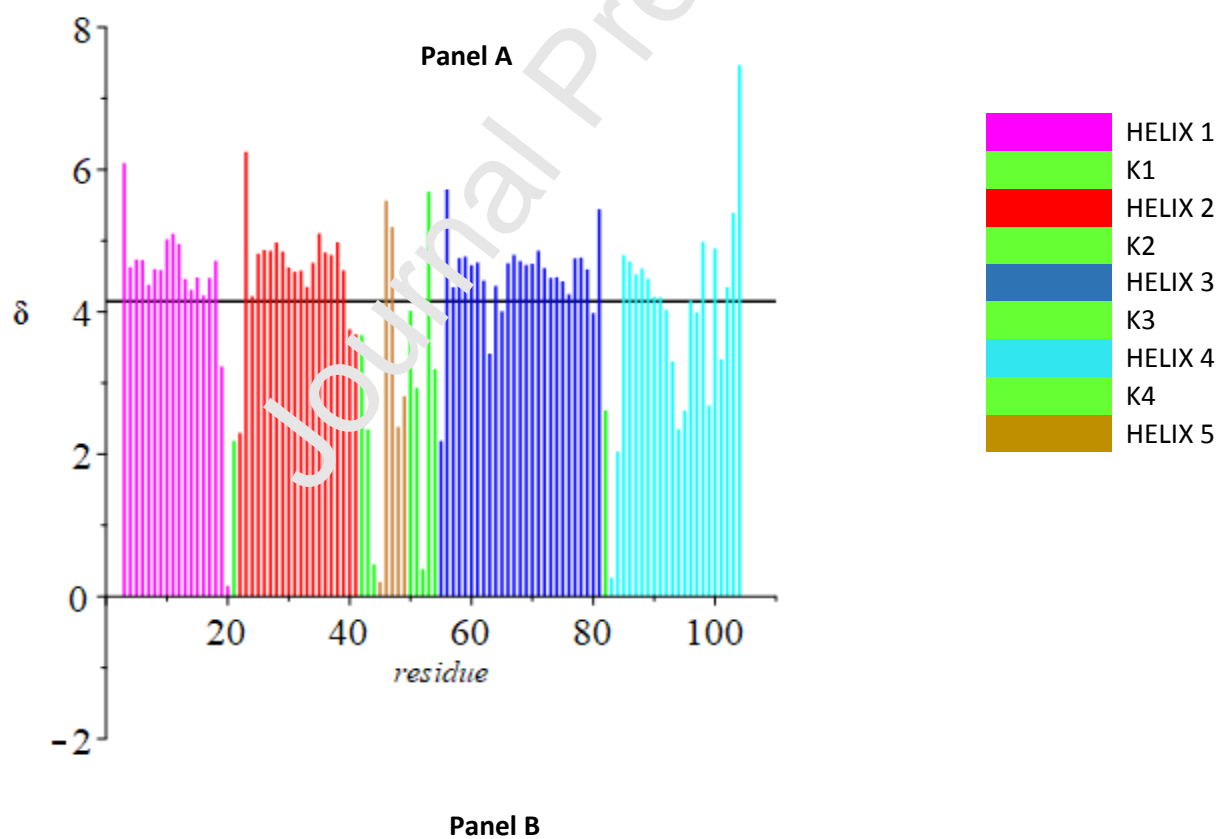
crystallographic distance between terminal α -carbons in the native state. For a five-residue segment centered on residue $i=8$,

$$ratio\ 08 = T08/R06to10.$$

Values of the *ratio* ~ 1 signal five-residue segments whose geometry is essentially that of the unfolded state; larger values of this ratio signal geometries of five-residue segments different from the unfolded state, reflecting the persistence of native-state regions at this stage of unfolding. Earlier we demonstrated the remarkable result that there is an exact correspondence between values of the *ratio* and the value f of the displacement of residue i from the origin (Fe) relative to the native state $f=1$ [16]. Hence, values of the *ratio* quantify *both* the persistence of the native-state geometry of the n -residue segment considered *and* the relative displacement of the central residue i in that segment from the helix.

Notice that both metrics depend only on relative distances and remain invariant regardless of the choice of origin of the coordinate system. These metrics are displayed for both proteins in Figures 4 and 5 for a five-residue segment.

Figure 4. Difference (δ in Å) of a 5-residue linear extension of triplets minus the distance between terminal α -carbons (i-2 to i+2). Upper /lower figure: cyt c-b₅₆₂/ Rd apo cyt b₅₆₂



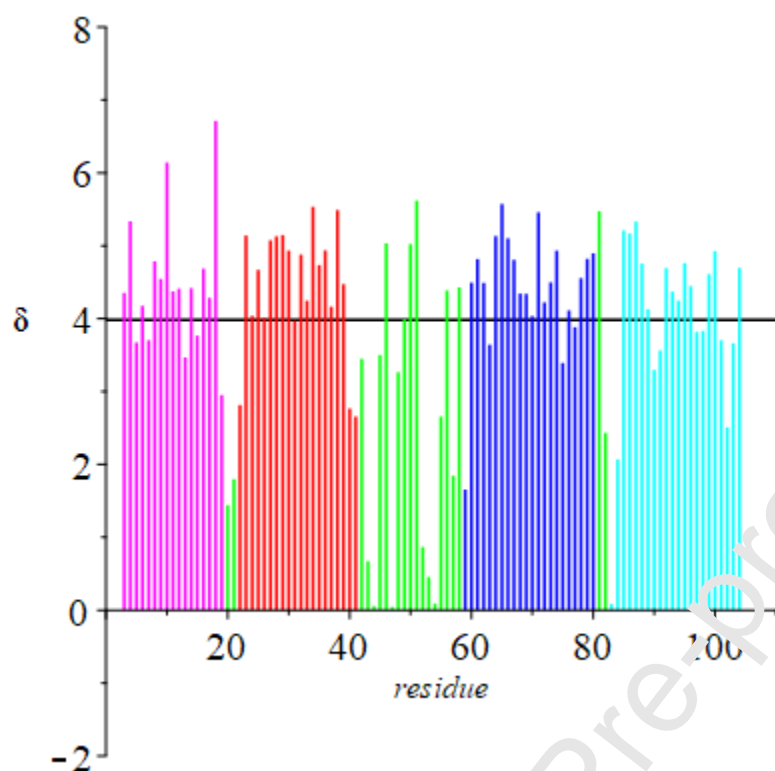
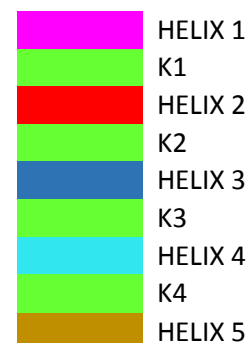
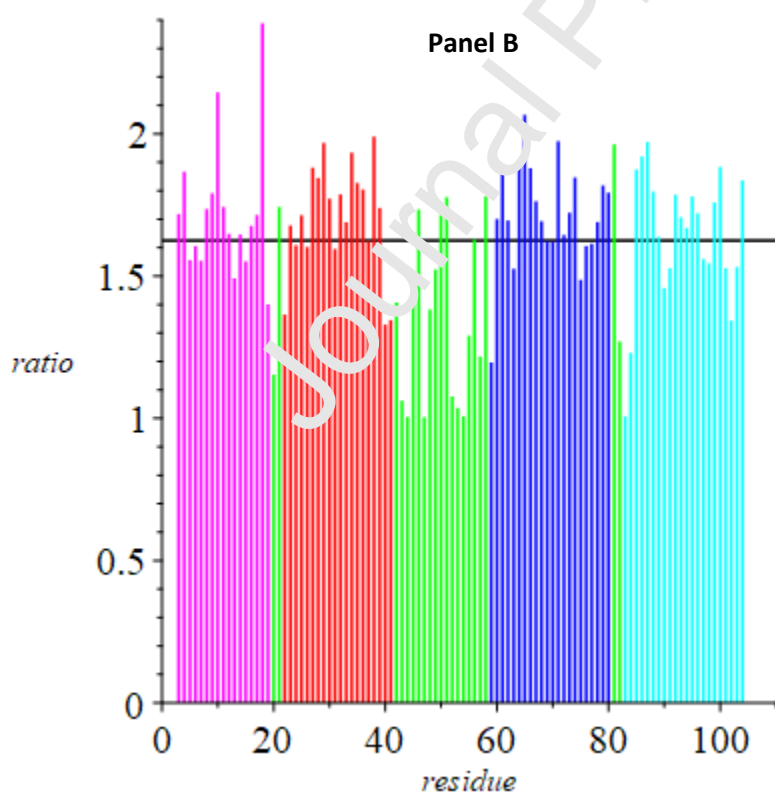
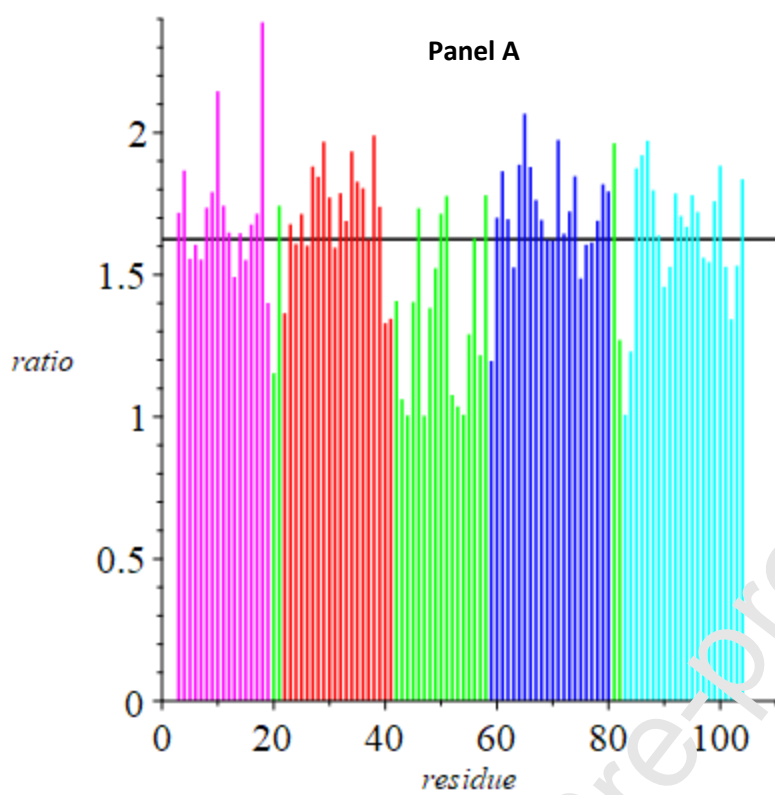


Figure 5. Ratio of a 5-residue linear extension of triplets to the distance between terminal α -carbons ($i-2$ to $i+2$). Upper /lower figure: cyt c-b₅₆₂/ Rd apo cyt b₅₆₂.





Two observations follow from an examination of the profiles displayed in Figures 4 and 5. First, whereas the profiles in the native state, Figure 3, for each region are sensibly “continuous,” those for the first unfolded state are not. Residues nearest turning regions have values of δ and *ratio* noticeably smaller than those in the interiors of helical regions. Second, there are residues in the turning regions that have values $\delta \sim 0$ and *ratio* values ~ 1 . These pivotal residues are identified in Table 1.

Table 1: Pivotal residues in turning regions. Data for a five-residue extension

Protein	Residue	Ratio	δ distance (Å)	Molecular Volume (Å ³)
cyt c-b ₅₆₂	ALA 20	1.014	0.150	88.6
	THR 44	1.035	0.453	116.1
	PRO 45	1.018	0.206	112.7
	SER 52	1.032	0.383	89.0
	LYS 83	1.024	0.262	168.6
Rd apo cyt b ₅₆₂	ALA 43	1.062	0.672	88.6
	THR 44	1.004	0.052	116.1
	LYS 47	1.004	0.043	168.6
	SER 47	1.077	0.867	89.0
	PRO 53	1.036	0.455	112.7
	ASP 54	1.008	0.092	111.1
	LYS 83	1.008	0.086	168.6

III. Helical Regions

The δ and *ratio* data displayed in Figures 4 and 5 for individual residues can be used to construct average values of those two metrics for each of the helical regions in cyt c-b₅₆₂ and the apo protein. Displayed in Tables A1 and A2 are values of δ for the helical regions for each of the six stages of unfolding; corresponding values of the *ratio* are in Tables A3 and A4.

As can be seen from the comparison presented in Table 2, at every extension except the first (segments of $n=5$ residues), the helix nearest the C-terminal end of the apo protein unfolds first. For the two interior helical regions, in every case the apo protein unfolds first. For the helix nearest the C-terminal end, except the first, the holo protein unfolds before the apo protein.

Based on this evidence for helices, we anticipate that the apo protein will be the first to denature. The situation when all residues are considered is taken up in the next section.

Table 2: Comparison of results for cyt c-b₅₆₂ and the apo protein derived from Tables 3-6. Holo helices in black and apo helices in red. At each extension, H or A denotes the more unfolded helix.

Extension (n)	H1,H1	H2,H2	H4,H3	H5,H5
5	H	A	A	A
7	A	A	A	H
9	A	A	A	H
11	A	A	A	H
13	A	A	A	H
15	A	A	A	H

IV. All-Residue Averages

The *overall* behavior of each protein at each stage of unfolding can be quantified by constructing the average over residues in both helical and turning regions. These averages are presented in Table 3 for S and Table 4 for the ratio. As forecast from the data in Table 2, the systematically smaller values for the apo protein for both metrics suggest that this protein unfolds first.

Extension	cyt c-b ₅₆₂	Rd apo cyt b ₅₆₂
5	4.148	3.983
7	6.432	6.270
9	9.754	9.389
11	13.02	12.54
13	16.48	15.90
15	20.29	19.76

Table 3: Difference δ (in Å) of n-residue linear extension of triplets minus the distance between terminal α -carbons. Average over *all* residues.

Table 4: Ratio of n-residue linear extension of triplets to the distance between terminal α -

Extension	cyt c-b ₅₆₂	Rd apo cyt b ₅₆₂
5	1.640	1.625
7	1.555	1.638
9	1.841	1.787
11	1.968	1.892
13	2.153	2.04
15	2.287	2.179

carbons. Average over *all* residues

V. Angular Phase Diagrams

In earlier work [14-16], we developed angular phase diagrams that provide an independent way of tracking the unfolding of a protein. As noted, the starting point in our approach is a triplet module of three residues, a center residue (i) flanked by its two first nearest neighbors $i-1$ and $i+1$. Adopting a coordinate system in which the iron atom is assigned as the origin, we use crystal structure data to calculate the distance $R(i-1)$ between the iron atom and the α -carbon of the left-most residue, the distance $R(i+1)$ to the right-most residue, and the distance $R(i-1)$ to $R(i+1)$ between the two α -carbons of the terminal residues. Also calculated from crystallographic

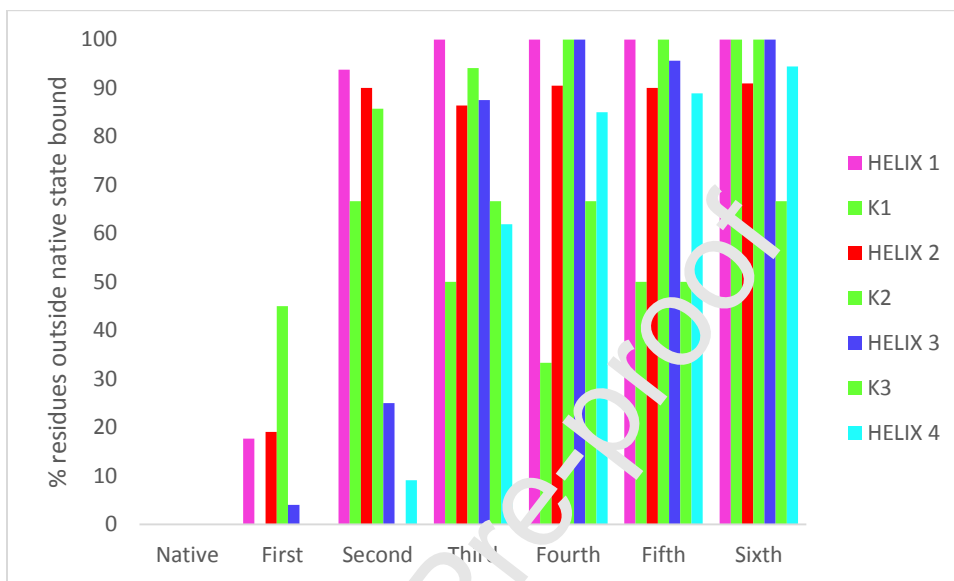
data are the angles between $R(i-1)$ and $R(i-1 \text{ to } i+1)$, $R(i-1 \text{ to } i+1)$, and $R(i-1 \text{ to } i+1)$ and $R(i+1)$, designated α , β , γ , respectively. These signatures were compiled for each of the n residues of the protein. Analogous calculations were carried out for sequences of five, seven, eleven, and fifteen residues [16].

As the protein unfolds from the native state, the angle β_i increases, whereas both angles α_i and γ_i *decrease* as the protein unfolds. Hence, the behavior of $[\alpha_i, \gamma_i]$ is reciprocal to the behavior of the two spatial metrics $[\delta, \text{ratio}]$, both of which *increase* as the protein denatures. When the protein begins to unfold, there is a reciprocal relation between angular and spatial descriptors.

Data for the spatial metric δ are given in Table A1. To make meaningful comparisons between different helices, average values that differ by $\leq 0.2 \text{ \AA}$ are regarded as the same. The data show that for the holo protein the native-state geometry of helix H2 near the N-terminal end of the polypeptide chain is *more* conserved than the terminal helix H1, while the angular phase-diagram data reported in [16] revealed that this is accompanied by a *greater* percent departure of angles $[\alpha_i, \gamma_i]$ from the native state for helix H1 than for H2.

This reciprocity also is displayed by the two helices H4 and H5 nearest the C-terminal end in the holo protein; the spatial descriptor δ documents that the native-state geometry of helix H4 is *more* conserved than the terminal helix H5, while the percent departure of angles $[\alpha_i, \gamma_i]$ from the native state is *greater* for H5 than for H4. The single exception is the last (sixth stage) of unfolding. The signature $[\alpha, \gamma]$ for the holo protein was given in [16]. This signature for Rd apo cyt b₅₆₂ is in Figure 6.

Figure 6. Percent departure of helical and non-helical regions from the native state $[\alpha, \gamma]$ domain for Rd apo cyt b_{562} .



Using the data for δ in concert with the $[\alpha, \gamma]$ signatures for Rd apo cyt b_{562} , we find *exactly* the same behavior for H1, H2 for the apo protein as for the holo protein, not a surprising result since the only residue different in the two proteins is residue 7 in the helix nearest the N-terminal end. However, for the two helices nearest the C-terminus, the similarity is not conclusive, as five residues are different in one helix (residues 98, 99, 101, 102 and 106).

VI. Discussion

The behavior of turning regions and the role of the Fe atom as the heme protein unfolds have been documented in this study. Displayed in Table 5 is the average distance difference δ and *ratio* in the extended turning region in Rd apo cyt b_{562} , residues 42-57. Compared with the helical regions in the apo protein, the $[\delta, \text{ratio}]$ values for this turning region are *smaller* in the first three stages of unfolding but *larger* in the last three stages.

Table 5: Average ratio and distance difference δ (in Å) in the extended non-helical (turning) region (residues 42-57) in Rd apo cyt b₅₆₂.

Extension (n)	Ratio	Difference Distance (Å)
5	1.329	2.560
7	1.494	5.550
9	1.615	8.741
11	1.805	12.62
13	1.990	16.92
15	2.294	21.29

Smaller values of these metrics signal that the turning region is more unfolded than helical regions, whereas larger values suggest that the turning region is less so. This “crossover” behavior is a consequence of the fact that H-bonds in the neighboring helices (H2, H3) are disrupted in going from the first stage of unfolding to the sixth stage. One turn of a helix involves 3.6 residues, two turns involve 7.2 residues, three turns 10.8 residues, and four turns 14.4 residues. The first stage of unfolding (5-residue segments) disrupts one turn of the helix, the third stage (9-residue segments) disrupts two turns, the fourth stage (11-residue segment) disrupts three turns and the sixth stage (15-residue segments) disrupts four turns. The crossover occurs when three turns of the helix are disrupted in the two helices bracketing the turning region.

The pivotal residues identified in Table 1 are turning points separating regions. The average molecular volume (Å³) of n-residue segments centered on each residue are given in Table 6. Values are reported for each stage n of unfolding, so neighborhoods capturing successively larger numbers of nearest neighbors are involved.

Table 6: Average molecular volume (\AA^3) of n-residue segments centered on residue i. Δ is the maximum difference of molecular volume values for all neighborhoods centered on residue i.

Protein	Residue i	n=5	n=7	n=9	n=11	n=13	n=15	$\Delta(\text{\AA}^3)$
cyt c-b ₅₆₂	ALA 20	88.6	122.8	124.2	125.2	128.8	133.8	45.2
	THR 44	116.1	105.8	119.7	130.2	128.7	130.1	24.4
	PRO 45	112.7	113.8	119.7	133.4	128.7	130.1	20.7
	SER 52	89.0	123.4	118.5	117.1	127.8	134.8	45.8
	LYS 83	168.6	122.9	135.1	132.6	128.7	128.6	45.7
Rd apo cyt c-b ₅₆₂	ALA 43	88.6	124.4	126.0	118.7	131.2	124.7	42.6
	THR 44	116.1	105.8	119.7	130.2	128.7	130.1	24.4
	LYS 47	168.6	149.3	139.8	132.3	131.0	126.5	42.1
	SER 52	89.0	123.4	118.5	117.1	127.8	130.9	41.9
	PRO 53	112.7	104.3	114.1	113.5	127.3	130.8	26.5
	ASP 54	111.1	104.3	102.9	117.4	127.5	135.9	33.0
	LYS 83	168.6	122.9	135.1	132.6	128.7	128.6	45.7

Based on a study of four heme proteins [cyt c-b₅₆₂, cyt c', sw-Mb, and h-Cygb] [16], we presented evidence that in the *de novo* synthesis of a turning region, residues are preferentially selected such that the range of molecular volume values (representing purely repulsive, excluded volume effects) span a small range of Δ values ($\Delta = 39.1 \text{ \AA}^3$) relative to the total range spanned ($\Delta = 167.7 \text{ \AA}^3$) by the 20 amino acids. In Table 6, Δ is the *maximum* difference of molecular volume values for first- through seventh-nearest neighbors centered on residue i. The overall average for cyt c-b₅₆₂ is 36.3 \AA^3 (36.6 \AA^3 for the apo protein). These data reinforce our conclusion that in the *de novo* synthesis of proteins [21], minimization of steric interactions in turning regions is of fundamental importance.

Results comparing the role of Fe in influencing cyt c-b₅₆₂ unfolding (*vs* the apo protein) are more nuanced [9, 22-30]. Tables A5 and A6 list values of the two metrics [δ , *ratio*] for residues within 10 \AA of the Fe atom for cyt c-b₅₆₂ and 10 \AA of the “virtual” Fe atom in the apo protein. Three

stages of unfolding are displayed, the first stage (5-residue segments), the third stage (9- residue segments), and the sixth stage (15-residue segments). For each residue i , the value of the metric for each protein is listed, pairwise, with the residue for the cytochrome with the larger value in red. Residue pairs for residue i that correspond to different residues in the two proteins are in blue. For metric δ , cyt c-b₅₆₂ in the first, third and sixth stage of unfolding has a greater number of residues with values greater than those of the apo protein: 18, 16 and 18 vs 16, 16 and 7. In compiling these data, two residues with essentially the same value were not included. For the *ratio* metric (cyt c-b₅₆₂) in the first, third and sixth stages, the numbers of residues having values greater than the apo protein are 16, 16, and 14, respectively. Those numbers for the apo protein are, respectively, 16, 12 and 10.

Our finding that three of the four helical regions in the apo protein unfold before corresponding helical regions in the holoprotein is not surprising, as heme iron ligation would be expected to stabilize the native folded structure [31], what is more, our model is in full accord with experimental investigations of thermal unfolding of apo and holo cytochrome b₅₆₂ [32, 33]. Turning to kinetics experiments, Bai *et al.* [20, 26-30] presented evidence that Rd apo cyt b₅₆₂ folds in an apparent two-state manner with the absence of detectable folding intermediates. In earlier protein engineering studies, they suggested that the rate-limiting reaction step (RLS) involves the formation of the two middle helices, H2 and H3. Based on the identification of three discrete intermediates, they proposed that Rd apo cyt b₅₆₂ folds with an early RLS. They noted that intermediates are formed only after the initial RLS and concluded that these states are hidden and hence that folding appears to be a two-step kinetics process. In our analysis, the $[\delta, \text{ratio}]$ and $[\alpha, \gamma]$ data show that H2>H1; that is, the geometry of the interior helix H2 is more conserved (*i.e.*, native like) than the exterior helix, H1. If the *last* helix to unfold is the *first* to

fold along the landscape leading to the native structure, then folding the interior helix is the RLS for the apo protein. And, as seems likely, cyt b₅₆₂ and cyt c-b₅₆₂ follow similar paths, the interior helix would protect the heme from misligation, thereby accounting for experimental results [9, 17, 18] suggesting that these folding energy landscapes are minimally frustrated.

Acknowledgments

We thank Devarajan (Dave) Thirumalai for very helpful comments. Work at Caltech was supported by the NIH (DK019038) and the Arnold and Mabel Beckman Foundation. Support at Pomona College was provided by the Howard Hughes Medical Institute Research Program and a Sontag Research Fellowship Award.

REFERENCES

1. Onuchic JN, Wolynes PG (2004) Theory of protein folding: Curr Opin Struct Biol 14, 70 – 75. a
2. Saven JG, Wolynes, PG (1996) Local conformational signals and the statistical thermodynamics of collapsed helical proteins: J Mol Biol 257, 199-216.
3. Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of Protein Folding: The energy landscape perspective: Ann Rev Phys Chem 48, 545-600.
4. Wolynes, PG (2015) Evolution, energy landscapes and the paradoxes of protein folding: Biochimie 119, 218-230.
5. Lin X., Schafer NP, Lu W, Jin, SX, Chen M., Onuchic JN, Wolynes, PG (2019) Forging tools for refining predicted protein structures: Proc Natl Acad Sci USA 116, 9400-9409.
6. Chiti F, Dodson CM (2017) Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade: Annu Rev Biochem 86, 27-68.
7. Chung HS, Eaton WA (2018) Protein folding transition path times from single molecule FRET: Curr Opin Struct Biol 46, 30-39.
8. Weinkam P, Pletneva EV, Gray HL, Winkler JR, Wolynes PG (2009) Electrostatic effects on funneled landscapes and structural diversity in denatured protein ensembles: Proc Natl Acad Sci USA 106, 1796 –1801.
9. Kimura T, Lee JC, Gray F.B, Winkler JR (2009) Folding energy landscape of cytochrome *cb*₅₆₂: Proc Natl Acad Sci USA 106, 7834 –7839.
10. Guo Z, Thirumalai D (1996) Kinetics and Thermodynamics of Folding of a de Novo Designed Four-helix Bundle Protein: J Mol Biol 263, 323-343.
11. Buchete N-V, Straub JE, Thirumalai D (2004) Development of novel statistical potentials for protein fold recognition: Curr Opin Struct Biol 14, 225-232.
12. Buchete N-V, Straub, JE, Thirumalai D (2003) Anisotropic coarse-grained statistical potentials improve the ability to identify nativelike protein structures: J Chem Phys 118, 7658-71.

13. Denesyuk NA, Thirumalai D (2015) How do metal ions direct ribozyme folding? *Nat Chem* 7, 793-801.
14. Kozak JJ, Gray HB (2019) Stereochemistry of residues in turning regions of helical proteins: *J Biol Inorg Chem* 24, 879-888.
15. Kozak JJ, Gray HB, Garza-Lopez RA (2016) Cytochrome unfolding pathways from computational analysis of crystal structures: *J Inorg Biochem* 155, 44-55.
16. Kozak JJ, Gray HB, Garza-Lopez RA (2020) Funneled angle landscapes for helical proteins: *J Inorg Biochem* 208, 111091.
17. Wittung-Stafshede P, Lee JC, Winkler JR, Gray HB (1999) Cytochrome b₅₆₂ folding triggered by electron transfer: Approaching the speed limit for formation of a four-helix-bundle protein: *Proc Natl Acad Sci USA* 96, 6587-6590.
18. Faraone-Mennella J, Gray HB, Winkler JR (2005) Early events in the folding of four-helix-bundle heme proteins: *Proc Natl Acad Sci USA* 102, 6315-6319.
19. Faraone-Mennella J, Tezcan FA, Gray HB, Winkler JR (2006) Stability and folding kinetics of structurally characterized cytochrome c-b₅₆₂: *Biochemistry* 45, 10504-10511.
20. Feng HQ, Takei J, Lipsitz R, Tjandra N, Bai YW (2003) Specific nonnative hydrophobic interactions in a hidden folding intermediate: implications for protein folding: *Biochemistry* 42, 12461-12467.
21. Korendovych IV, DeGrado WF (2020) De novo Protein Design, a Retrospective. *Quart Rev Biophys* 53, E3.
22. Lin Y-W, Ni F-Y, Yang T-L (2009) Early events in thermal unfolding of apocytochrome b₅₆₂ and its double-cysteine mutant as revealed by molecular dynamics simulation: *J Mol Struct* 898, 82-89.
23. Manyasa S, Whitford D (1999) Defining folding and unfolding reactions of apocytochrome b₅ using equilibrium and kinetic fluorescence measurements: *Biochemistry* 38, 9533-9540.
24. Wang T, et al. (2007) Probing the folding intermediate of Rd-apocytochrome b₅₆₂ by Protein engineering and infrared T-jump: *Protein Sci* 16, 1176-1183.
25. Wang LJ, Sun N, Terzyan S, Zhang XJ, Benson DR (2006) Histidine/tryptophan π -stacking interaction stabilizes the heme-independent folding core of microsomal

- apocytochrome b5 relative to that of mitochondrial apocytochrome b5: *Biochemistry* 45, 13750-13759.
26. Feng, HQ Vu ND, Zhou Z, Bai, YW (2004) Structural examination of ϕ -value analysis in Protein folding: *Biochemistry* 43, 14325-14331.
 27. Choy WY, Zhou Z, Bai YW, Kay LE (2005) An ^{15}N NMR spin relaxation dispersion study of the folding of a pair of engineered mutants of apocytochrome b₅₆₂: *J Am Chem Soc* 127, 5066-5072.
 28. Feng H, Zhou Z, Bai YW (2005) A protein folding pathway with multiple folding Intermediates at atomic resolution: *Proc Natl Acad Sci USA* 102, 5026-5031.
 29. Zhou Z, Huang YZ, Bai YW (2005) An on-pathway hidden intermediate and the early rate-limiting transition state of Rd-apocytochrome b₅₆₂ characterized by protein engineering: *J Mol Biol* 352, 757-764.
 30. Feng H, Takei J, Lipsitz N, Bai Y (2003) Specific non-native hydrophobic interactions in a hidden folding intermediate. Implications for protein folding: *Biochemistry* 42 12461-12465.
 31. Tezcan, FA, Winkler JR, Gray HB (1998) Effects of Ligation and Folding on Reduction Potentials of Heme Proteins: *J Am Chem Soc* 120, 13383-13388.
 32. Robinson CR, Liu Y, O'Brien P, Sligar SG, Sturtevant JM (1998) A differential scanning calorimetric study of the thermal unfolding of apo- and holo-cytochrome b₅₆₂: *Protein Science* 7, 961-965.
 33. Robinson, CR, Liu, Y, Thomson, JA, Sturtevant, JM, Stephen G. Sligar, SG (1997) Energetics of Heme Binding to Native and Denatured States of Cytochrome b₅₆₂: *Biochemistry* 36, 16141-16146.

APPENDIX

Table A1: Difference δ (in Å) of n-residue linear extension of triplets minus the distance between terminal α -carbons for cyt c-b₅₆₂. Average for helical regions (residues in parentheses).

Extension (n)	Helix 1 (2-20)	Helix 2 (22-41)	Helix 3 (45-49)	Helix 4 (55-81)	Helix 5 (83-106)
5	4.382	4.584	3.231	4.501	3.971
7	6.512	6.740	6.811	6.596	5.881
9	9.509	9.933	11.12	9.676	8.819
11	12.50	12.96	16.21	12.70	11.58
13	16.19	16.20	20.67	16.04	14.36
15	19.91	19.90	25.40	19.37	17.91

Table A2: Difference δ (in Å) of n-residue linear extension of triplets minus the distance between terminal α -carbons for Rd apo cyt b₅₆₂. Average for helical regions (residues in parentheses).

Extension (n)	Helix 1 (1-19)	Helix 2 (22-41)	Helix 3 (58-80)	Helix 4 (55-81)
5	4.459	4.441	4.419	3.995
7	6.228	6.634	6.320	6.219
9	8.424	8.809	9.379	9.095
11	11.00	12.72	12.13	12.10
13	13.78	16.14	15.07	14.91
15	15.81	19.68	18.39	18.72

Table A3: Ratio of n-residue linear extensions of triplets to the distance between terminal α -carbons for cyt c-b₅₆₂. Average for helical regions (residues in parentheses).

Extension (n)	Helix 1 (2-20)	Helix 2 (22-41)	Helix 3 (45-49)	Helix 4 (55-81)	Helix 5 (83-106)
5	1.697	1.725	1.466	1.723	1.640
7	1.651	1.709	1.650	1.714	1.579
9	1.768	1.830	1.939	1.818	1.781
11	1.845	1.940	2.292	1.906	1.809
13	2.075	2.141	2.476	2.137	1.817

15	2.183	2.238	2.586	2.136	2.062
----	-------	-------	-------	-------	-------

Table A4: Ratio of n-residue linear extensions of triplets to the distance between terminal α -carbons for Rd apo cyt b₅₆₂. Average for helical regions (residues in parentheses).

Extension (n)	Helix 1 (1-19)	Helix 2 (22-41)	Helix 3 (58-60)	Helix 4 (83-106)
5	1.720	1.705	1.717	1.640
7	1.623	1.705	1.650	1.638
9	1.706	1.831	1.784	1.793
11	1.769	1.928	1.830	1.857
13	1.882	2.152	1.897	1.932
15	1.993	2.238	2.005	2.067

Table A5. Difference δ (in Å) in distance of n-residue linear extension of triplets minus the distance between terminal alpha carbons for residues within 10 Å of Fe. Red denotes larger value in pair.

Residue i	n=5		n=9		n=15	
	cyt c-b ₅₆₂	Rd apo cyt c-b ₅₆₂				
LEU 3	6.091	4.354				
GLU 4	4.628	5.337				
ASP 5	4.733	3.674	12.46	8.143		
ASN 6	4.728	4.174	9.291	10.12		
MET 7	4.379		9.330			
TRP 7		3.704		8.273		
GLU 8	4.602	4.786	9.769	10.35	18.87	15.01
THR 9	4.589	4.547	9.508	8.211	16.83	18.66
LEU 10	5.020	6.144	9.618	9.580	16.88	14.24
ASN 11	5.098	4.374	9.145	9.153	16.62	17.36
ASP 12	4.953	4.413	9.405	10.57	16.60	15.36
ALA 36	4.835	4.937	9.700	11.08	16.58	16.43
ALA 37	4.801	4.164	9.751	9.248	16.03	17.88
LEU 38	4.977	5.489	8.711	8.265	15.99	14.91
ASP 39	4.582	4.475	8.618	7.759	17.04	16.23
ALA 40	3.757	2.766	3.738	9.087	16.31	16.09
GLN 41	3.683	2.658	9.253	8.176	17.74	17.62
LYS 42	3.674	3.450	6.571	4.815	16.71	13.30
ALA 43	2.348	0.672	4.959	6.297	16.86	15.33
THR 44	0.453	0.052	10.41	9.242	18.62	16.50
PRO 45	0.2064	3.501	9.169	7.250	22.01	17.73
PRO 46	5.5627	5.035	8.592	7.980	25.04	19.76
GLY 64	4.365	5.134	9.253	9.926	15.98	14.84
PHE 65	4.006	5.572	8.244	8.701	16.72	16.30
LEU 94	2.352	4.248	8.237	9.195	13.77	17.90
LYS 95	2.611	4.761	7.686	8.487	16.17	16.14
THR 96	4.160	4.445	7.717	8.262	13.11	14.45
THR 97	3.993	3.818	5.544	9.401	15.21	16.10
CYS 98	4.984		9.137		14.77	
ILE 98		3.832		9.419		17.53
ASN 99	2.685		7.536		19.04	
ARG 99		4.611		7.913		16.35

ALA 100	4.894	4.927	9.385	7.236
CYS 101	3.332		8.097	
HIS 101		3.705		8.879
HIS 102	4.345		12.84	
ASN 102		2.510		9.653

Journal Pre-proof

Table A6. Ratio of n-residue linear extension of triplets to the distance between terminal alpha carbons for residues within 10 Å of Fe. Red denotes larger value.

Residue i	n=5		n=9		n=15	
	cyt c-b ₅₆₂	Rd apo cyt c-b ₅₆₂	106	apo 106	106	apo 106
LEU 3	1.981	1.718				
GLU 4	1.730	1.867				
ASP 5	1.773	1.556	2.190	1.643		
ASN 6	1.759	1.606	1.735	1.800		
MET 7	1.700		1.758			
TRP 7		1.553		1.673		
GLU 8	1.723	1.736	1.806	1.847	1.917	1.699
THR 9	1.740	1.791	1.775	1.680	1.788	1.887
LEU 10	1.850	2.146	1.781	1.741	1.797	1.646
ASN 11	1.850	1.743	1.723	1.650	1.774	1.775
ASP 12	1.825	1.649	1.765	1.867	1.778	1.733
ALA 36	1.774	1.804	1.798	1.970	1.768	1.805
ALA 37	1.794	1.623	1.607	1.788	1.683	1.815
LEU 38	1.831	1.990	1.662	1.591	1.695	1.645
ASP 39	1.738	1.740	1.650	1.577	1.721	1.683
ALA 40	1.534	1.329	1.605	1.655	1.713	1.723
GLN 41	1.502	1.345	1.690	1.587	1.772	1.765
LYS 42	1.430	1.467	1.371	1.261	1.733	1.525
ALA 43	1.247	1.052	1.278	1.388	1.712	1.614
THR 44	1.035	1.004	1.755	1.634	1.884	1.717
PRO 45	1.018	1.404	1.649	1.471	2.110	1.742
PRO 46	1.866	1.734	1.550	1.502	2.587	1.945
GLY 64	1.657	1.888	1.728	1.832	1.742	1.655
PHE 65	1.619	2.067	1.627	1.670	1.787	1.762
LEU 94	1.280	1.669	1.618	1.761	1.622	1.922
LYS 95	1.326	1.779	1.643	1.668	1.739	1.776
THR 96	1.617	1.721	1.545	1.626	1.576	1.638
THR 97	1.828	1.560	1.390	1.773	1.659	1.785
CYS 98	1.812		1.705		1.682	
ILE 98		1.545		1.805		1.889

ASN 99	1.417		1.618		1.941	
ARG 99		1.758		1.589		1.801
ALA 100	1.778	1.883	1.729	1.538		
CYS 101	1.440					
HIS 101		1.528				
HIS 102	1.641					
ASN 102		1.343				

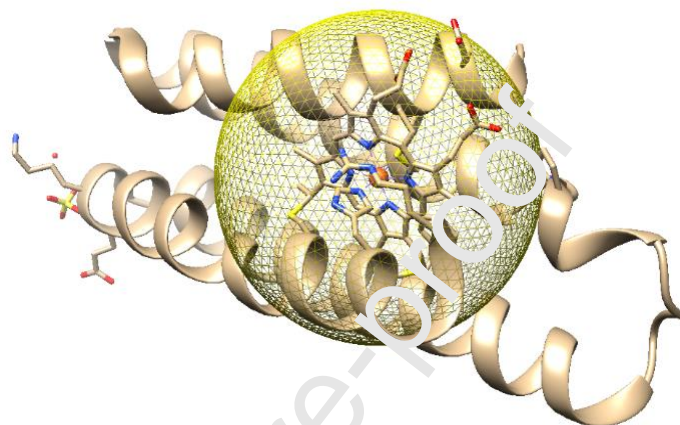
Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

GRAPHICAL ABSTRACT (SYNOPSIS)



The folding landscape of cytochrome c-b₅₆₂ is minimally frustrated, owing to helix encapsulation of the heme.

HIGHLIGHTS

- Geometrical model to study unfolding of two cytochromes.
- Model quantifies the resiliency of the native state to steric perturbations.
- Development of angular phase diagrams.
- Sequential unfolding of helical regions is predicted.
- Percent departure of helical and non-helical regions is calculated.

Journal Pre-proof